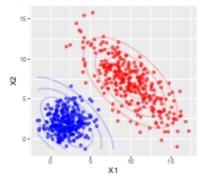
LINEAR DISCRIMINANT ANALYSIS (LDA) /2

LDA assumes that each class density is modeled as a *multivariate* Gaussian:

$$p(\mathbf{x}|y=k) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu_k)^T \Sigma^{-1}(\mathbf{x}-\mu_k)\right)$$

with equal covariance, i. e. $\Sigma_k = \Sigma \quad \forall k$.





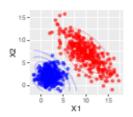
LINEAR DISCRIMINANT ANALYSIS (LDA) /3

Parameters heta are estimated in a straightforward manner by estimating

$$\hat{\pi_k} = \frac{n_k}{n}$$
, where n_k is the number of class- k observations

$$\hat{\mu_k} = \frac{1}{n_k} \sum_{i: y^{(i)} = k} \mathbf{x}^{(i)}$$

$$\hat{\Sigma} = \frac{1}{n-g} \sum_{k=1}^{g} \sum_{i: \gamma^{(i)} = k} (\mathbf{x}^{(i)} - \hat{\mu_k}) (\mathbf{x}^{(i)} - \hat{\mu_k})^T$$





LDA AS LINEAR CLASSIFIER / 2

We can formally show that LDA is a linear classifier, by showing that the posterior probabilities can be written as linear scoring functions - up to any isotonic / rank-preserving transformation.

$$\pi_k(\mathbf{x}) = \frac{\pi_k \cdot \rho(\mathbf{x}|y=k)}{\rho(\mathbf{x})} = \frac{\pi_k \cdot \rho(\mathbf{x}|y=k)}{\sum\limits_{j=1}^g \pi_j \cdot \rho(\mathbf{x}|y=j)}$$

As the denominator is the same for all classes we only need to consider

$$\pi_k \cdot p(\mathbf{x}|y=k)$$

and show that this can be written as a linear function of x.



LDA AS LINEAR CLASSIFIER / 3

$$\pi_{k} \cdot p(\mathbf{x}|y = k)$$

$$\propto \qquad \pi_{k} \exp\left(-\frac{1}{2}\mathbf{x}^{T} \boldsymbol{\Sigma}^{-1} \mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_{k}^{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{k} + \mathbf{x}^{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{k}\right)$$

$$= \exp\left(\log \pi_{k} - \frac{1}{2}\boldsymbol{\mu}_{k}^{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{k} + \mathbf{x}^{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{k}\right) \exp\left(-\frac{1}{2}\mathbf{x}^{T} \boldsymbol{\Sigma}^{-1} \mathbf{x}\right)$$

$$= \qquad \exp\left(\theta_{0k} + \mathbf{x}^{T} \boldsymbol{\theta}_{k}\right) \exp\left(-\frac{1}{2}\mathbf{x}^{T} \boldsymbol{\Sigma}^{-1} \mathbf{x}\right)$$

$$\propto \qquad \exp\left(\theta_{0k} + \mathbf{x}^{T} \boldsymbol{\theta}_{k}\right)$$

by defining
$$\theta_{0k} := \log \pi_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k$$
 and $\theta_k := \Sigma^{-1} \mu_k$.

We have again left out all constants which are the same for all classes k, so the normalizing constant of our Gaussians and $\exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right)$.

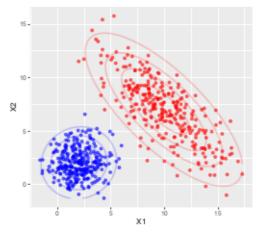
By finally taking the log, we can write our transformed scores as linear:

$$f_k(\mathbf{x}) = \boldsymbol{\theta}_{0k} + \mathbf{x}^T \boldsymbol{\theta}_k$$



QUADRATIC DISCRIMINANT ANALYSIS (QDA) /2

- Covariance matrices can differ over classes.
- Yields better data fit but also requires estimation of more parameters.





QUADRATIC DISCRIMINANT ANALYSIS (QDA) /3

$$\begin{array}{lcl} \pi_k(\mathbf{x}) & \propto & \pi_k \cdot p(\mathbf{x}|y=k) \\ & \propto & \pi_k |\Sigma_k|^{-\frac{1}{2}} \exp(-\frac{1}{2}\mathbf{x}^T \Sigma_k^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_k^T \Sigma_k^{-1}\boldsymbol{\mu}_k + \mathbf{x}^T \Sigma_k^{-1}\boldsymbol{\mu}_k) \end{array}$$

Taking the log of the above, we can define a discriminant function that is quadratic in x.

$$\log \pi_k - \frac{1}{2} \log |\Sigma_k| - \frac{1}{2} \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k + \mathbf{x}^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k - \frac{1}{2} \mathbf{x}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{x}$$



QUADRATIC DISCRIMINANT ANALYSIS (QDA) /4

